Euphytica (2009) 169:285–301
DOI 10.1007/s10681-009-9917-4

Using molecular markers and field performance data to characterize the Pee Dee cotton germplasm resources

B. T. Campbell · V. E. Williams · W. Park

Received: 18 December 2008 / Accepted: 2 March 2009 / Published online: 13 March 2009

Abstract Knowledge of genetic relationships in crop breeding programs provides valuable information that can be used by plant breeders as a parental line selection tool. In Upland cotton (*Gossypium hirsutum* L.), the Pee Dee germplasm program represents one of the most historically significant Upland cotton breeding programs and is known as a key source of fiber quality genes for commercial cultivars. The foundation of the Pee Dee germplasm is known to represent an array of genetic diversity involving the hybridization of *G. hirsutum* L., *G. barbadense* L., and triple hybrid strains (*G. arboreum* L. × *G. thurberi* Todaro × *G. hirsutum* L.). In this study, we characterized genetic relationships within the Pee Dee germplasm collection using molecular marker and field performance data. Molecular marker and field performance data showed the Pee Dee germplasm collection still maintains useful amounts of genetic diversity. The methods described provide plant breeders of cotton and other crops a strategy to develop a parental line selection tool based on genotypic and phenotypic information. Cotton breeders can directly use the information provided to select specific Pee Dee germplasm parental line combinations based on genotypic (molecular marker) and phenotypic (field performance) information rather than relying on pedigree and phenotypic information alone.

Keywords Cotton · Genetic diversity · Fiber quality · Lint yield · Plant breeding

Abbreviations

AFIS	Advanced fiber information system
HVI	High volume instrument
UPGMA	Unweighted pair group method with arithmetic average
SSR	Simple sequence repeat

Introduction

Assessing the genetic relationships in crop breeding programs provides valuable information that can be exploited by plant breeders. Plant breeders can use the knowledge of genetic relationships among breeding lines, germplasm lines, and cultivars as a parental line selection tool. It is widely accepted that genetic diversity must exist between parental line combinations to develop recombinant segregating populations representing new and favorable combinations. Hence, a large number of genetic diversity studies have been conducted on a wide assortment of diverse plant genera using a variety of methodologies. A few recent,

B. T. Campbell (✉) · V. E. Williams · W. Park
USDA-ARS Coastal Plains Soil, Water and Plant Research Center, 2611 W. Lucas St., Florence, SC 29501, USA
e-mail: todd.campbell@ars.usda.gov

Present Address:
V. E. Williams
Department of Microbiology and Immunology, Medical University of South Carolina, Charleston, SC 29425, USA

but not all inclusive, examples of genetic diversity studies in a number of plants include wheat (*Triticum aestivum* L.; Fufa et al. 2005), maize (*Zea mays* L.; Labate et al. 2003), rice (*Oryza sativa* L.; Garris et al. 2005), pearl millet (*Pennisetum glaucum* L.; Budak et al. 2003), soybean [*Glycine max* (L.) Merr] (Brown-Guedira et al. 2000), and peanut (*Arachis hypogaea* L.; Milla et al. 2005). Mohammadi and Prasanna (2003) provide an informative review of the methodologies most often used to analyze genetic diversity in different types of populations.

Over the past 20 years, there have been numerous reports exploring the genetic diversity of the Upland cotton (*Gossypium hirsutum* L.) primary gene pool. Early reports assessing genetic relationships among cotton genotypes were based on pedigree relationships (coefficient of parentage) (May et al. 1995; Van Esbroeck et al. 1998, 1999), morphological and phenotypic traits (Van Esbroeck et al. 1999), isozyme markers (Wendel et al. 1992), and restriction fragment length polymorphism (RFLP) markers (Brubaker and Wendel 1994; Van Becelaere et al. 2005). More recently, other types of molecular markers have been used to assess genetic relationships among specific cotton genotypes. These include random amplified polymorphic DNA (RAPD) markers (Tatineni et al. 1996; Khan et al. 2000; Lu and Myers 2002; Rahman et al. 2002), amplified fragment length polymorphism (AFLP) markers (Pillay and Myers 1999; Abdalla et al. 2001; Iqbal et al. 2001), and microsatellite or simple sequence repeat (SSR) markers (Liu et al. 2000; Gutierrez et al. 2002; Zhang et al. 2005a, b; Lacape et al. 2007). Several of these studies also compared one or more method of estimating genetic relationships among a group of cotton genotypes including a comparison of coefficient of parentage to phenotypic trait analysis (Van Esbroeck et al. 1999), coefficient of parentage to RFLP (Van Becelaere et al. 2005), and morphological trait analysis to RAPD (Tatineni et al. 1996).

The early studies of genetic diversity within cotton, based on pedigree relationships and coefficient of parentage, reported extensive genetic diversity among the various germplasm groups representing the primary gene pool of Upland cotton (May et al. 1995; Van Esbroeck et al. 1998). However, Van Esbroeck et al. (1999) and Van Becelaere et al. (2005) subsequently reported the likelihood that pedigree based coefficient of parentage genetic similarity estimates were inflated because of numerous false assumptions used to calculate pedigree based coefficient of parentage. Recent studies using molecular markers suggest a fairly high degree of genetic uniformity and similarity. Van Becelaere et al. (2005) and Lu and Myers (2002) reported very high levels of genetic similarity ranging from 0.91 to 0.97 and 0.93 to 0.98, respectively. Abdalla et al. (2001) reported a mean genetic similarity of 0.86 and Rahman et al. (2002) reported a range of 0.82–0.95. Each of these studies surveyed a number of historically important US Upland cotton germplasm lines and cultivars except for Rahman et al. (2002) that surveyed genotypes from Pakistan. Other studies present a broader range of genetic similarity among cotton genotypes. Gutierrez et al. (2002) reported a genetic similarity range of 0.66–0.94 when analyzing a set of US Upland cotton cultivars. Zhang et al. (2005a, b) reported genetic similarities ranging from 0.69 to 0.94 and 0.62 to 0.94 in two separate reports focused on US Upland cotton cultivars and New Mexico Acala cotton germplasm, respectively.

A portion of the Upland cotton primary gene pool resources has been described by Bowman et al. (2006), where they provide pedigree information along with a historical description of 13 different US cotton breeding and germplasm development programs. Two of the historically relevant US cotton breeding and germplasm development programs highlighted by Bowman et al. (2006) are the New Mexico Acala and Pee Dee germplasm programs. Bowman and Gutierrez (2003) estimated the New Mexico Acala and Pee Dee germplasm programs accounted for over 50% of fiber strength improvements present in commercial cultivars. The remaining fiber strength improvements were attributed to transgressive segregants of unknown origin.

Although the exact origin of the fiber strength genes present in the New Mexico Acala and Pee Dee germplasm programs has not been determined experimentally, the common breeding history of both germplasm programs suggests interspecific introgression. Both germplasm programs were initiated by hybridizing germplasm from *G. barbadense* L., the Triple Hybrid (*G. arboreum* L. × *G. thurberi* Todaro × *G. hirsutum* L.), and *G. hirsutum* L. (Smith et al. 1999). Since the wide genetic foundation of each germplasm program provided unique opportunities for recombination among the genomes of two tetraploid species (AD) and two diploid species (A and D), it is plausi-

ble that each germplasm program maintains considerable levels of genetic diversity. Zhang et al. (2005b) measured the genetic diversity among the New Mexico Acala germplasm collection using SSR markers and reported a genetic similarity ranging from 0.62 to 0.94 with a mean similarity of 0.82. They concluded the unique foundation of the germplasm program resulted in high levels of genetic diversity resulting from gene introgression.

The foundation and history of the Pee Dee germplasm program is quite complex and was nicely described by Culp and Harrell (1973). Briefly, the Pee Dee germplasm program was initiated in 1935 to improve the yield and boll weevil (*Anthonomous grandis* Boheman) tolerance of Sea Island cotton (*G. barbadense* L.) and to develop early-maturing, high-yielding, extra-long staple (ELS) Upland cottons with Sea Island fiber properties. By the mid 1940s, two ELS Upland cultivars (Sealand and Earlistaple) were developed and commercially grown on limited hectares. At this time, breeding objectives of the Pee Dee germplasm program changed and focused on developing Upland cultivars with the combination of high strength and yield. Around the same time, unique triple hybrid strains (*G. arboreum* L. × *G. thurberi* Todaro × *G. hirsutum* L.) with high fiber strength were developed and distributed to cotton breeding programs (Beasley 1940; Kerr 1960). Triple hybrid strains TH 108 and TH 171 were crossed to a series of Upland cotton parents including 'Hopi Acala', Sealand, and Earlistaple to develop the F, J, A, and N Upland progenitor lines. Over the next 50 years, intercrossing and crossing strategies among the progenitor lines and with commercial cultivars resulted in the development of a wealth of Upland cultivars and germplasm lines. Although Culp and Harrell (1973) attributed early success in the Pee Dee germplasm program to pedigree selection with early generation testing, they also described their use of alternative breeding methods such as random intermating, modified backcrossing, and composite crossing. As Hanson described in 1959 (Hanson 1959), random intermating should provide the maximum opportunity to break up linkage blocks using repetitive recombination followed by selfing. Hence, it is plausible that the unique foundation of the Pee Dee germplasm program and alternative breeding methods employed would develop a genetically broad array of Upland cotton germplasm.

Considering the unique history of the Pee Dee germplasm program, our hypothesis is that considerable genetic diversity exists within the Pee Dee germplasm. In this study, our objective was to characterize the genetic relationships within the Pee Dee germplasm collection using molecular marker and phenotypic field performance data.

Materials and methods

Plant materials and DNA extraction

Eighty-two officially released cotton germplasm lines and/or cultivars were selected from the Pee Dee cotton germplasm collection to represent the history of the Pee Dee cotton breeding program. Care was taken to select a range of lines representative of different pedigrees and points in time over the life of the breeding program. This list of germplasm lines and cultivars is provided in Table 1. Five plants of each genotype were grown in a greenhouse to obtain leaf tissue. After approximately six weeks, a bulk leaf tissue sample was obtained from the five plants of each genotype and used for DNA extraction. Two grams of leaf tissue were ground using a mortar and pestle and liquid nitrogen. Genomic DNA was extracted from each genotype following the procedure of Lu and Myers (2002). The extracted DNA was re-suspended in 500 µl TE buffer and the DNA concentration was quantified using a DNA fluorometer (GE Healthcare). Subsequently, DNA concentrations for each genotype were adjusted to a standard concentration of 1 µg μl^{-1}.

Molecular marker analysis

We used a two-step strategy to select molecular markers for genetic diversity assessment. The first step was to identify cotton simple sequence repeat (SSR) markers identified as polymorphic in previous studies identified in the Cotton Marker Database (Blenda et al. 2006). From the pool of polymorphic markers identified, the second step was to select markers to ensure complete coverage of each cotton linkage group or chromosome. Overall, our goal was to assay 2 markers per chromosome arm for a total of 104 molecular markers to provide a genome wide survey of genetic diversity. Primers for each marker were

Table 1 Least square means and cluster analysis grouping for Pee Dee lines and commercial check genotypes evaluated for lint percent, lint yield, fiber length, fiber strength, and fiber fineness from 2004 to 2006 near Florence, SC

Genotype	Lint percent		Lint yield		Fiber length		Fiber strength		Fiber fineness		Registration reference
	%	Group	kg ha^{-1}	Group	mm	Group	kN m kg^{-1}	Group	mg km^{-1}	Group	
AC235(9)	37.1	1.1	1,151	1.1	29.7	1	303.3	1.1	174.9	1.1	Culp and Harrell (1980a)
AC241	36.7	3.1	1,117	1.1	28.8	2.1	308.1	1.3	173.6	1.1	Culp and Harrell (1980a)
Earlistaple7	38.5	2.1	1,052	1.1	29.0	2.2	302.8	1.1	179.4	2.1	Culp and Harrell (1980c)
FJA347	35.7	3.1	1,010	1.2	30.5	1	295.9	1.2	171.2	1.2	Culp and Harrell (1980c)
FTA266	36.1	3.1	1,080	1.1	30.3	1	304.4	1.1	171.5	1.2	Culp and Harrell (1980c)
Hybrid330-278	34.9	3.2	995	1.2	30.7	1	318.9	1.3	171.5	1.2	Culp and Harrell (1980c)
PD-1	39.8	2.3	1,222	1.3	29.1	2.2	288.4	2.1	178.7	2.1	Culp et al. (1985a)
PD-2	38.3	1.2	1,046	1.1	28.1	2.3	272.9	2.2	176.6	1.1	Culp et al. (1985b)
PD-3	39.3	2.2	1,241	1.3	28.7	2.1	291.4	2.1	173.6	1.1	Culp et al. (1988)
PD-3-14	39.6	2.3	1,439	2.1	29.2	2.2	296.0	1.2	175.7	1.1	May et al. (1996)
PD0109	39.0	2.2	1,196	1.3	29.2	2.2	301.0	1.1	178.2	2.1	Culp and Harrell (1980b)
PD0111	38.1	1.2	1,129	1.1	28.8	2.1	285.1	2.1	175.2	1.1	Culp and Harrell (1980b)
PD0113	37.4	1.1	1,141	1.1	29.2	2.2	291.0	2.1	173.8	1.1	Culp and Harrell (1980b)
PD0259	37.6	1.1	1,216	1.3	28.9	2.1	288.7	2.1	173.0	1.2	Harrell and Culp (1979a)
PD0648	39.1	2.2	1,246	1.3	28.2	2.3	273.0	2.2	174.8	1.1	Culp et al. (1990a)
PD0683	37.4	1.1	1,101	1.1	28.9	2.2	284.4	2.1	173.3	1.1	Culp et al. (1990a)
PD0723	38.7	2.1	1,192	1.3	28.2	2.3	271.3	2.2	174.5	1.1	Culp et al. (1990a)
PD0738	37.7	1.1	1,125	1.1	29.3	2.2	292.8	1.2	173.5	1.1	Culp et al. (1990b)
PD0741	38.6	2.1	1,179	1.3	28.9	2.1	280.0	2.3	178.3	2.2	Culp et al. (1990b)
PD0747	39.1	2.2	1,202	1.3	28.0	2.3	279.1	2.3	178.0	2.2	Culp et al. (1990b)
PD0753	38.8	2.1	1,111	1.1	28.3	2.3	271.6	2.2	180.1	2.1	Culp et al. (1990b)
PD0756	39.2	2.2	1,257	1.4	28.4	2.3	278.7	2.3	179.6	2.1	Culp et al. (1990b)
PD0761	38.6	2.1	1,367	2.2	28.4	2.3	282.0	2.3	177.3	2.2	Culp et al. (1990b)
PD0762	38.5	2.1	1,146	1.1	28.6	2.1	280.0	2.3	176.0	1.1	Culp et al. (1990b)
PD0771	38.5	2.1	1,209	1.3	28.6	2.1	277.6	2.3	176.4	1.1	Culp et al. (1990b)
PD0778	39.0	2.2	1,261	1.4	28.6	2.1	279.9	2.3	174.1	1.1	Culp et al. (1990b)
PD0781	39.2	2.2	1,145	1.1	27.5	2.4	276.0	2.3	179.4	2.1	Culp et al. (1990b)
PD0785	38.9	2.1	1,278	1.4	27.7	2.4	276.1	2.3	182.1	2.1	Culp et al. (1990b)
PD0804	38.7	2.1	1,081	1.1	28.1	2.3	275.1	2.3	181.7	2.1	Culp et al. (1990b)
PD0878	38.0	1.2	1,186	1.3	28.4	2.3	277.1	2.3	181.0	2.1	Culp et al. (1990a)
PD0948	38.8	2.1	1,265	1.4	28.8	2.1	284.2	2.1	175.8	1.1	Culp et al. (1990a)
PD2164	38.5	2.1	1,097	1.1	29.1	2.2	311.7	1.3	173.7	1.1	Culp and Harrell (1980a)
PD2165	34.9	3.2	976	1.2	30.0	1	298.4	1.2	184.6	3	Harrell and Culp (1979a)
PD3246	37.8	1.1	1,032	1.1	30.1	1	301.6	1.1	176.0	1.1	Culp and Harrell (1980a)
PD3249	38.1	1.2	1,199	1.3	29.0	2.2	288.3	2.1	172.4	1.2	Culp and Harrell (1980a)
PD4381	37.8	1.1	1,074	1.1	29.1	2.2	297.4	1.2	174.6	1.1	Harrell and Culp (1979b)
PD4461	38.7	2.1	1,165	1.3	28.7	2.1	287.3	2.1	174.6	1.1	Culp and Harrell (1979c)
PD5246	39.0	2.2	1,387	2.2	28.8	2.1	288.0	2.1	179.5	2.1	Green et al. (1991b)
PD5256	38.5	2.1	1,435	2.1	29.1	2.2	298.1	1.2	180.2	2.1	Green et al. (1991b)
PD5286	39.2	2.2	1,372	2.2	28.7	2.1	290.1	2.1	176.5	1.1	Green et al. (1991a)
PD5358	39.5	2.3	1,265	1.4	29.2	2.2	298.8	1.2	174.7	1.1	Green et al. (1991b)
PD5363	38.8	2.1	1,342	2.2	29.6	1	305.1	1.1	178.2	2.2	Green et al. (1991c)
PD5377	38.7	2.1	1,387	2.2	29.1	2.2	295.3	1.2	175.6	1.1	Green et al. (1991b)
PD5380	39.5	2.3	1,177	1.3	28.8	2.1	288.0	2.1	175.2	1.1	Green et al. (1991b)
PD5472	39.6	2.3	1,333	2.2	28.7	2.1	280.8	2.3	177.9	2.2	Green et al. (1991c)

Table 1 continued

Genotype	Lint percent		Lint yield		Fiber length		Fiber strength		Fiber fineness		Registration reference
	%	Group	kg ha^{-1}	Group	mm	Group	kN m kg^{-1}	Group	mg km^{-1}	Group	
PD5529	39.0	2.2	1,268	1.4	29.6	1	295.8	1.2	172.7	1.2	Green et al. (1991a)
PD5576	38.7	2.1	1,354	2.2	28.5	2.3	278.7	2.3	174.7	1.1	Green et al. (1991a)
PD5582	39.0	2.2	1,189	1.3	28.6	2.1	289.6	2.1	178.8	2.2	Green et al. (1991a)
PD6044	38.7	2.1	1,123	1.1	28.7	2.1	293.9	1.2	174.7	1.1	Culp et al. (1985c)
PD6132	38.6	2.1	1,070	1.1	29.1	2.2	301.8	1.1	175.7	1.1	Culp et al. (1985c)
PD6142	38.2	1.2	1,236	1.3	29.2	2.2	293.8	1.2	171.4	1.2	Culp et al. (1985c)
PD6179	38.3	1.2	1,273	1.4	29.0	2.2	287.7	2.1	172.3	1.2	Culp et al. (1985c)
PD6186	38.2	1.2	1,266	1.4	29.0	2.2	306.6	1.3	175.1	1.1	Culp et al. (1985c)
PD695	36.7	3.1	1,092	1.1	29.3	2.2	281.6	2.3	172.3	1.2	Harrell and Culp (1979b)
PD6992	37.8	1.1	1,270	1.4	29.6	1	297.9	1.2	177.1	2.2	Culp et al. (1985c)
PD7388	38.0	1.2	1,338	2.2	28.8	2.1	284.8	2.1	181.4	2.1	Culp et al. (1990a)
PD7439	37.7	1.1	1,240	1.3	28.8	2.1	290.1	2.1	176.7	1.1	Culp et al. (1990a)
PD7458	38.2	1.2	1,013	1.2	28.3	2.3	271.1	2.2	173.4	1.1	Culp et al. (1990a)
PD7496	37.5	1.1	1,109	1.1	29.3	2.2	282.6	2.3	176.8	1.1	Culp et al. (1990a)
PD7501	37.4	1.1	1,226	1.3	28.1	2.3	284.6	2.1	174.2	1.1	Culp et al. (1990a)
PD7586	38.0	1.2	1,268	1.4	28.5	2.3	280.5	2.3	178.6	2.2	Culp et al. (1990a)
PD7723	39.3	2.2	1,179	1.3	28.8	2.1	279.8	2.3	184.2	3	Culp et al. (1990a)
PD8619	37.4	1.1	1,183	1.3	28.9	2.2	295.0	1.2	169.1	1.2	Culp and Harrell (1979d)
PD875	40.1	2.3	1,121	1.1	28.3	2.3	277.7	2.3	179.2	2.1	Harrell and Culp (1979b)
PD9223	39.4	2.2	1,165	1.3	28.4	2.3	281.8	2.3	178.5	2.2	Culp and Harrell (1979b)
PD9232	38.1	1.2	1,174	1.3	29.0	2.2	297.4	1.2	175.9	1.1	Culp and Harrell (1979b)
PD93007	39.9	2.3	1,152	1.1	28.5	2.3	286.3	2.1	180.1	2.1	May and Howle (1997b)
PD93009	39.1	2.2	1,283	1.4	28.4	2.3	290.4	2.1	179.9	2.1	May and Howle (1997a)
PD93019	39.8	2.3	1,284	1.4	28.3	2.3	285.2	2.1	177.1	2.2	May and Howle (1997a)
PD93021	39.6	2.3	1,363	2.2	28.9	2.1	294.2	1.2	177.4	2.2	May and Howle (1997a)
PD93030	39.9	2.3	1,319	2.2	28.7	2.1	292.7	1.2	179.8	2.1	May and Howle (1997a)
PD93034	39.5	2.3	1,150	1.1	29.0	2.2	293.5	1.2	177.7	2.2	May and Howle (1997a)
PD93043	38.7	2.1	1,183	1.3	29.2	2.2	292.5	1.2	181.6	2.1	May and Howle (1997b)
PD93046	39.0	2.2	1,331	2.2	28.6	2.1	286.5	2.1	176.4	1.1	May and Howle (1997b)
PD93057	38.8	2.1	1,172	1.3	28.9	2.1	283.7	2.1	177.7	2.2	May and Howle (1997a)
PD9363	38.6	2.1	1,442	2.1	29.3	2.2	300.4	1.1	180.9	2.1	Culp and Harrell (1979b)
PD9364	39.0	2.2	1,392	2.2	29.3	2.2	291.0	2.1	180.2	2.1	Culp and Harrell (1979b)
PD94042	41.4	No cluster	1,361	2.2	28.8	2.1	289.9	2.1	182.5	2.1	May (1999)
PD94045	38.9	2.1	1,297	1.4	28.3	2.3	285.3	2.1	180.8	2.1	May (2001)
SC-1	38.8	2.1	1,361	2.2	28.8	2.1	294.9	1.2	176.0	1.1	Culp and Harrell (1979a)
Sealand542	34.8	3.2	885	No cluster	31.8	No cluster	312.4	1.3	163.6	No cluster	Culp and Harrell (1980c)
DP491	42.3		1,475		29.9		287.0		177.6		
DP444BR	41.8		1,254		27.6		263.3		181.2		
DP555BR	42.6		1,612		27.9		262.5		181.1		
FM958	41.5		1,346		28.9		295.1		185.6		
FM960BR	39.6		1,437		28.1		305.2		189.9		
ST5599BR	40.9		1,668		28.5		282.4		193.9		
Mean	38.6		1,184		28.8		289.1		176.6		
P-value[a]	<0.0001		0.0010		<0.0001		<0.0001		<0.0001		
LSD (0.05)	0.9		161		0.5		8.8		4.5		

[a] Probability associated with the *F*-test for differences among gentoypes using analysis of variance

synthesized using published sequence information obtained from the Cotton Marker Database. The polymerase chain reaction (PCR) mix preparation, PCR conditions, and gel electrophoresis were performed as described by Nguyen et al. (2004). Annealing temperature varied from 45 to 52°C depending on the T_m of each individual marker's primer set. PCR products were electrophoresed through 12% polyacrylamide gels at 200 V for 2 h. Following electrophoresis, gels were stained with 1% ethidium bromide for 30 min, destained in ddH_2O for 1 h, and visualized under UV light using an Alpha Imager gel documentation system (Alpha Innotech).

Phenotypic evaluations

The 82 Pee Dee genotypes and 2–6 commercial cultivars were evaluated in replicated field trials from 2004 to 2006 at Florence, SC to obtain estimates of agronomic performance and fiber quality. The 84–88 total genotypes were randomly assigned to a single replicate of a 4-replicate, α-lattice incomplete block field design for each trial. The α-lattice in 2004 consisted of 14 incomplete blocks of size 6, while in 2005 and 2006 there were 8 incomplete blocks of size 11. In 2004, a total of 84 genotypes included 'Deltapine 491' (DP491) and 'FiberMax 958' (FM958) as commercial checks. In 2005 and 2006, a total of 88 genotypes was evaluated, which consisted of the 82 Pee Dee lines, the previous 2 commercial checks, and four additional commercial checks that included 'Deltapine 555 BR' (DP555BR), 'Deltapine 444 BR' (DP444BR), 'Stoneville 5599 BR' (ST5599BR), and 'FiberMax 960 BR' (FM960BR). Each entry was grown in a two-row plot 16.5 m long with 76 cm spacing between rows, and trial management followed the established local practices.

Each plot was harvested with a spindle type mechanical cotton picker that harvested both rows of each plot, and total seed cotton weight was recorded. A 25-boll sample was obtained from each plot prior to harvest to determine lint percent and fiber quality properties. Samples were ginned on a 10-saw laboratory gin and lint percent was determined by dividing the weight of the lint sample after ginning by the weight of the lint sample before ginning. Lint yield was calculated by multiplying the lint percent by the seed cotton yield. In addition to lint yield and lint percent, a portion of the lint sample was sent to the Cotton Incorporated Fiber Testing Laboratory (Cary, NC) for determination of High Volume Instrument (HVI) and Advanced Fiber Information System (AFIS) fiber properties. The fiber properties selected and recorded for each plot included HVI fiber length, HVI fiber strength, and AFIS fineness. Analysis of variance combined across environments was conducted using the PROC GLM module of SAS ver. 9.2 (SAS Institute 2008) to test for differences among genotypes for each trait. Least square means of each genotype (adjusted for experimental design) were calculated for each of the traits measured. The least significant difference (LSD, $P = 0.05$) was calculated and used to provide a general comparison of genotypes for each trait. Pearson correlations were calculated using PROC CORR for each phenotypic trait combination to identify relationships between phenotypic traits.

Genetic diversity estimation

Amplified fragments of each SSR marker-genotype combination were scored as "1" and "0", where "1" and "0" indicated the presence or absence of a specific allele (band), respectively. In addition, the least squares adjusted genotype means for lint percent, lint yield, fiber length, fiber strength, and fineness were used to calculate average taxonomic distances between genotypes for each trait independently. All genetic diversity analyses were conducted using NTSYSpc ver.2.2 (Rohlf 2005). For molecular markers, genetic similarities between pairs of genotypes were measured using the SIMQUAL module by the DICE similarity coefficient based on the proportion of shared alleles (Dice 1945; Nei and Li 1979). For the phenotypic trait data, the DIST option of the SIMINT module was used to measure the average taxonomic distances between genotypes. The SAHN module was used to cluster genotypes based on the unweighted pair group method with arithmetic average (UPGMA) algorithm that calculates a similarity matrix for molecular markers and a dissimilarity matrix for phenotypic traits. Following, the TREE module was used to construct a dendrogram for each cluster analysis. For the molecular marker analysis, mean similarities were calculated within and between each cluster and sub-cluster using the similarity matrix. For the phenotypic trait analysis, mean taxonomic distances were calculated within and between each cluster and sub-cluster using the dissimilarity matrix. Clusters and sub-clusters

were identified within the dendrogram for molecular markers and each phenotypic trait following the method used by Brown-Guedira et al. (2000). For the molecular markers, clusters and sub-clusters were acceptable when the mean similarity among two or more genotypes within a cluster was greater than the overall mean similarity and the between cluster similarity was less than the within cluster similarity of either cluster or sub-cluster being compared. For the phenotypic traits, clusters and sub-clusters were acceptable when the mean dissimilarity among two or more genotypes within a cluster was less than the overall mean dissimilarity and the between cluster dissimilarity was greater than the within cluster dissimilarity of either cluster or sub-cluster being compared.

The DICE genetic similarity matrix calculated from the molecular marker data was compared to the average taxonomic distance matrices calculated for each of the phenotypic traits. In addition, all pairwise combinations of average taxonomic distance matrices for each trait were compared. All matrix comparisons were carried out using the MXCOMP module of NTSYSpc ver.2.2 (Rohlf 2005) that uses a normalized Mantel *Z* statistic. Beer et al. (1993) provides statistical details related to the Mantel *Z* statistic.

Results and discussion

Molecular marker analysis and diversity

Initially, we attempted to genotype the 82 Pee Dee lines with 115 SSR markers. Eighty of the SSR markers amplified reproducible bands, while the remaining 35 SSR markers did not amplify well or generated non-reproducible bands. Following, the 80 reproducible SSR markers were used to genotype 77 of the 82 Pee Dee lines. Five of the Pee Dee lines were discarded from the molecular marker analysis because of a high incidence of missing data. Out of the 80 SSR markers, 40% produced polymorphic bands across the Pee Dee lines. A total of 336 bands were produced from the 80 SSR markers for an average of 4.2 bands per marker. Of the 336 total bands produced, 20% were polymorphic across the Pee Dee lines. The 80 SSR markers and 336 total bands provided coverage of the 26 Upland cotton chromosomes with at least 1 marker per chromosome. The relatively low level of polymorphism found in this study corresponds to other reports indicating a low level of polymorphism within Upland cotton (Zhang et al. 2005b).

To calculate genetic similarities among lines, we selected a subset of the 80 SSR markers to ensure genome coverage and line discrimination ability. First, we attempted to select polymorphic markers (if available) for each of the 26 linkage groups. Seventeen of the linkage groups were covered by one or more polymorphic markers. However, from our SSR marker survey of the Pee Dee lines, we were unable to find polymorphic markers for 9 of the 26 linkage groups. To ensure marker coverage of the remaining 9 linkage groups, we included a single, monomorphic locus for each linkage group in the genetic similarity analysis. In total, we included 81 bands amplified from 41 markers for a genome-wide genetic diversity estimate. Each band was considered a single genetic locus. Table 2 provides a listing and description of the 41 SSR markers used to calculate estimates of genetic similarity. Chromosome assignments presented in Table 2 follow the Upland cotton nomenclature provided by Wang et al. (2006).

Results of the genetic similarity and cluster analysis are presented in the form of a dendrogram in Fig. 1. Overall, the Pee Dee lines clustered into 2 primary groups. However, PD0259, PD7388, AC-241, and PD0747 did not cluster with any of the other Pee Dee lines. Based on the cluster analysis and dendrogram, PD0259 and PD0747 represented the most diverse pairwise combination of lines with a genetic similarity of 0.64. The first primary group contained 49 lines and could be further separated into 7 subgroups of lines. In primary group 1, the pairwise genetic similarity ranged from 0.74 to 0.96. The second primary group contained 22 lines and could be further separated into 4 sub-groups of lines. In primary group 2, the pairwise genetic similarity ranged from 0.73 to 0.94. The genetic similarity level separating primary groups 1 and 2 was 0.71. PD-2 and PD7439 clustered together with a genetic similarity of 0.78. PD7388 and AC241 did not readily cluster with either of the 2 primary groups, but were more similar to the 2 primary groups than PD0259 and PD0747. The range of genetic similarity found within this study (0.64–0.96) is comparable to previous studies (Gutierrez et al. 2002; Rahman et al. 2002; Zhang et al. 2005a, b), but lower than the genetic similarities reported by Van Becelaere et al. (2005) and Lu and

Table 2 SSR markers used to estimate genetic similarities among Pee Dee germplasm lines

SSR marker	Chromosome	Polymorphism	SSR marker	Chromosome	Polymorphism	SSR marker	Chromosome	Polymorphism
NAU1067	A1	No	NAU1028	D3	Yes	NAU1009	A9	No
CIR234	D1	Yes	NAU869	A4	No	BNL1672	D9	Yes
CIR311	D1	No	CIR172	D4	Yes	BNL1414	D9	Yes
CIR015	D1	Yes	NAU1042	A5, D5	Yes	JESPR110	D9	Yes
JESPR101	A2	Yes	CIR253	A5	Yes	NAU921	A10	Yes
NAU895	A2	No	NAU1221	D5	Yes	NAU1297	D10	No
CIR288	D2	No	CIR165	D5	Yes	NAU1162	A11	Yes
NAU1070	A3	Yes	CIR179	A6	No	CIR196	A11	Yes
NAU1081	A3	Yes	NAU905	D6	Yes	NAU1366	D11	No
NAU1190	A3	Yes	CIR407	D6	Yes	CIR293	A12	Yes
NAU1248	A3	Yes	NAU1048	A7	Yes	NAU1231	D12	No
NAU862	A3	Yes	CIR107	D7	Yes	NAU1215	A13	Yes
CIR263	A3	Yes	NAU1369	A8	Yes	CIR099	D13	No
CIR212	A3, D5	Yes	NAU1322	D8	Yes			

Myers (2002). Interestingly, the range of genetic similarity within this study is very similar to that reported by Zhang et al. (2005b). Since the Pee Dee and New Mexico Acala germplasm programs had similar, genetically broad foundations, it is not surprising their similar range in genetic similarity. The higher genetic similarity reported by Van Becelaere et al. (2005) and Lu and Myers (2002) likely reflects the germplasm sampled.

Most sister line Pee Dee genotypes (lines with same pedigree) clustered into the same primary group, although there were a few exceptions. PD0778 did not cluster into primary group 1 as the 11 other sister lines did. Three other pairs of sister lines, PD7496 and PD7501, PD9223 and PD9232, and AC241 and AC235(9), also did not cluster into the same primary group. Interestingly, PD-3 and PD-3-14 did not cluster into the same primary group even though PD-3-14 was developed from a re-selection of PD-3. When examining the earliest progenitors of the Pee Dee germplasm collection included in this study, it was interesting that Earlistaple7, Hybrid330-278, FTA266, and FJA347 all clustered into primary group 1, while Sealand542 clustered into primary group 2. This cluster pattern is not surprising since Earlistaple7 was included in the complex crossing and selection system that gave rise to Hybrid330-278, FTA266, and FJA347. It is surprising that Sealand542 clustered into primary group 2, even though Sealand542 and Earlistaple7 share the common parent Coker Wilds. This suggests that Hybrid330-278, FTA266, and FJA347 retained a greater number of Earlistaple7 alleles as opposed to Sealand542 alleles during their development. This would support the idea that Sealand542 contains a relative larger number of *G. barbadense* alleles that are not easily introgressed into *G. hirsutum* backgrounds using traditional hybridization. Limited *G. barbadense* introgression has been documented in numerous reports, and recent molecular evidence further supports this claim (Jiang et al. 2000). The first germplasm lines produced from the initial progenitors of the Pee Dee germplasm collection mostly clustered in primary group 1 and included PD4461, PD3246, PD3249, AC235(9), PD4381, PD2165, and PD2164. The exceptions to this pattern are PD0259 and AC241 that did not cluster with any of the Pee Dee material. In addition, PD8619 did not follow this pattern and clustered within primary group 2.

Overall, the results of the molecular marker analysis revealed moderate to high levels of genetic similarity among the Pee Dee germplasm. This is not unexpected since the predominant breeding method used over the last 50 years has been inbreeding and pedigree selection within crosses primarily involving recycled and new combinations of Pee Dee lines (Culp and Harrell 1973). However, this analysis also provides evidence that useful genetic variability still exists within the Pee Dee germplasm. The preponderance of genetic variability still existing in the Pee germplasm may have

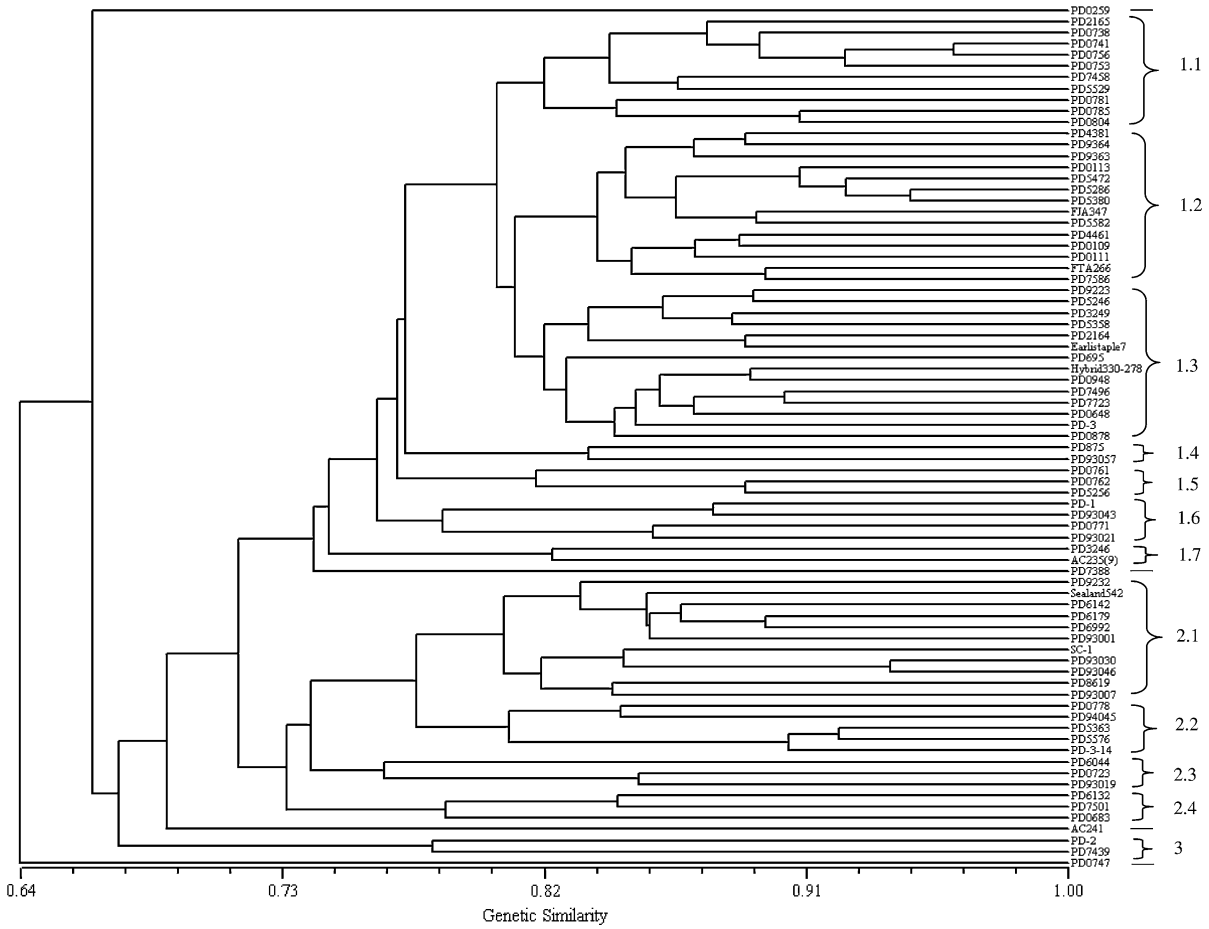

Fig. 1 Dendrogram representing the cluster analysis of Pee Dee germplasm *lines* based on molecular markers

resulted from the alternative breeding methods, in addition to pedigree selection, used within the first 25 years of the breeding program. Culp and Harrell (1973) described random intermating, backcrossing, and composite crossing methods that were used in addition to pedigree selection as approaches to simultaneously improve yield and fiber quality. They reported the success of both modified backcrossing and composite crossing methods to improve yield and fiber quality. Although they initially reported the failure of their random intermating method to produce any desirable recombinants at the time of publication (1973), it is possible this procedure began favorable recombination events that were not capitalized upon until many years later after pure lines were selected from these initial random mating populations. According to the early report by Hanson (1959), such a scheme should theoretically maintain and increase the genetic diversity of a germplasm pool if employed.

The ability to separate these Pee Dee lines using genome-wide SSR markers provides important information to the cotton breeding community. The cluster analysis and dendrogram shows that pedigree information is not always representative of true genetic diversity. Other molecular based genetic diversity studies have also demonstrated deficiencies related to measuring genetic diversity based on pedigree information only (Van Becelaere et al. 2005). The dendrogram presented here can be used as a parental line selection tool if a breeder is interested in selecting parental lines based on genetic dissimilarity.

Phenotypic trait analysis and diversity

In addition to examining molecular based genetic relationships, we examined agronomic and fiber quality performance relationships among the Pee Dee germplasm collection. Table 1 shows the least

squares adjusted mean performance of Pee Dee lines for lint percent, lint yield, and several fiber quality properties. Analysis of variance indicated significant differences among genotypes for each of the traits. When comparing Pee Dee lines against the best performing check for each trait, 9% and 21% of the Pee Dee lines performed equivalent to the highest check for fiber length and fiber strength, respectively. Hybrid330-278 outperformed the highest check (FM960BR) for fiber strength. Comparing the Pee Dee lines against the best performing check for fiber fineness, 14% of the Pee Dee lines outperformed the most fine check cultivar DP491. None of the Pee Dee lines performed equivalent to or better than the highest performing check for lint percent and lint yield. When comparing Pee Dee lines against the lowest performing commercial check, 41 (51%) and 64 (79%) performed equivalent to the lowest check for lint percent and lint yield, respectively. PD94042 out performed the lowest lint percent check (FM960BR), while PD5256 and PD9363 outperformed the lowest yielding check (DP491). For fiber length, 7% of the Pee Dee lines performed equivalent to and 93% better than the lowest fiber length check (DP444BR). For fiber strength, 2% of the Pee Dee lines performed equivalent to and 98% better than the lowest fiber strength check (DP555BR). For fiber fineness, 100% of the Pee Dee lines performed better than the least fine fiber fineness check (ST5599BR).

Pearson correlation analyses identified unfavorable and favorable relationships among the phenotypic traits measured in this study. Unfavorable correlations were found involving pairwise combinations of lint percent and fiber length ($r = -0.22$, $P < 0.01$), lint percent and fiber strength ($r = -0.18$, $P < 0.01$), and fiber strength and fiber fineness ($r = 0.40$, $P < 0.01$). Although the lint percent correlations support the well documented negative relationships previously reported for cotton fiber properties and lint percent (Culp and Harrell 1973; Meredith 2005), the lower magnitude of their correlation coefficients suggests the negative relationships are not as strong within the Pee Dee germplasm surveyed in this study as reported for other groups of cotton genotypes. The unfavorable relationship between fiber strength and fiber fineness indicates that selecting Pee Dee lines with strong fibers may inadvertently also select for coarser fibers. The favorable correlations identified for pairwise combinations of lint percent and lint yield ($r = 0.23$, $P < 0.01$), lint percent and fiber fineness ($r = -0.21$, $P < 0.01$), and fiber length and fiber strength ($r = 0.54$, $P < 0.01$) provide evidence that each trait combination can be favorably improved. Correlations involving lint yield with fiber length and fiber fineness were not significant and the correlation involving lint yield and fiber strength was very low ($r = -0.08$, $P < 0.05$). Lint yield correlations indicate that selecting high yielding Pee Dee lines should not negatively impact fiber properties per se.

Using least squares adjusted genotype means for each trait, we employed cluster analysis to separate or group lines based on trait performance. Results of similarity and cluster analyses for each trait are presented in the form of dendrograms in Figs. 2, 3, 4, 5, and 6. Table 1 also provides the group and sub-group clustering designation of genotypes for each trait. Instead of using principal component or other similar analyses to group lines based on information for all traits simultaneously, traits were each considered independently to avoid any effects of autocorrelation or scaling issues among traits. Figure 2 displays the dendrogram for lint percent. The Pee Dee lines separated into 3 primary groups, with 23 lines present in primary group 1, 50 lines in primary group 2, and only 7 lines in primary group 3. PD94042 did not readily cluster with any of the primary groups but was more similar to groups 1 and 2 rather than 3. Primary group 1 could be further separated into 2 sub-groups, primary group 2 into 3 sub-groups, and primary group 3 into 2 sub-groups. The mean lint percent for primary group 1 was 37.8%, 39.1% for group 2, and 35.7% for group 3. Four of the early progenitor Pee Dee lines clustered into group 3 with the lowest mean lint percent. The remaining progenitor line Earlistaple7 clustered into sub-group 2.1 which produced an average mean lint percent.

Figure 3 displays the dendrogram for lint yield. The Pee Dee lines separated into 2 primary groups, with 63 lines present in primary group 1 and 17 in primary group 2. Sealand542 did not readily cluster with any of the other Pee Dee lines. Primary group 1 could be further clustered into 4 sub-groups and primary group 2 could be separated into 2 additional sub-groups. The mean lint yield for primary group 1 was 1,167 kg ha^{-1} and 1,372 kg ha^{-1} for primary group 2. All of the early, progenitor genotypes of the Pee Dee germplasm, excluding Sealand542, clustered into sub-groups 1.1 and 1.2, representing the lowest mean

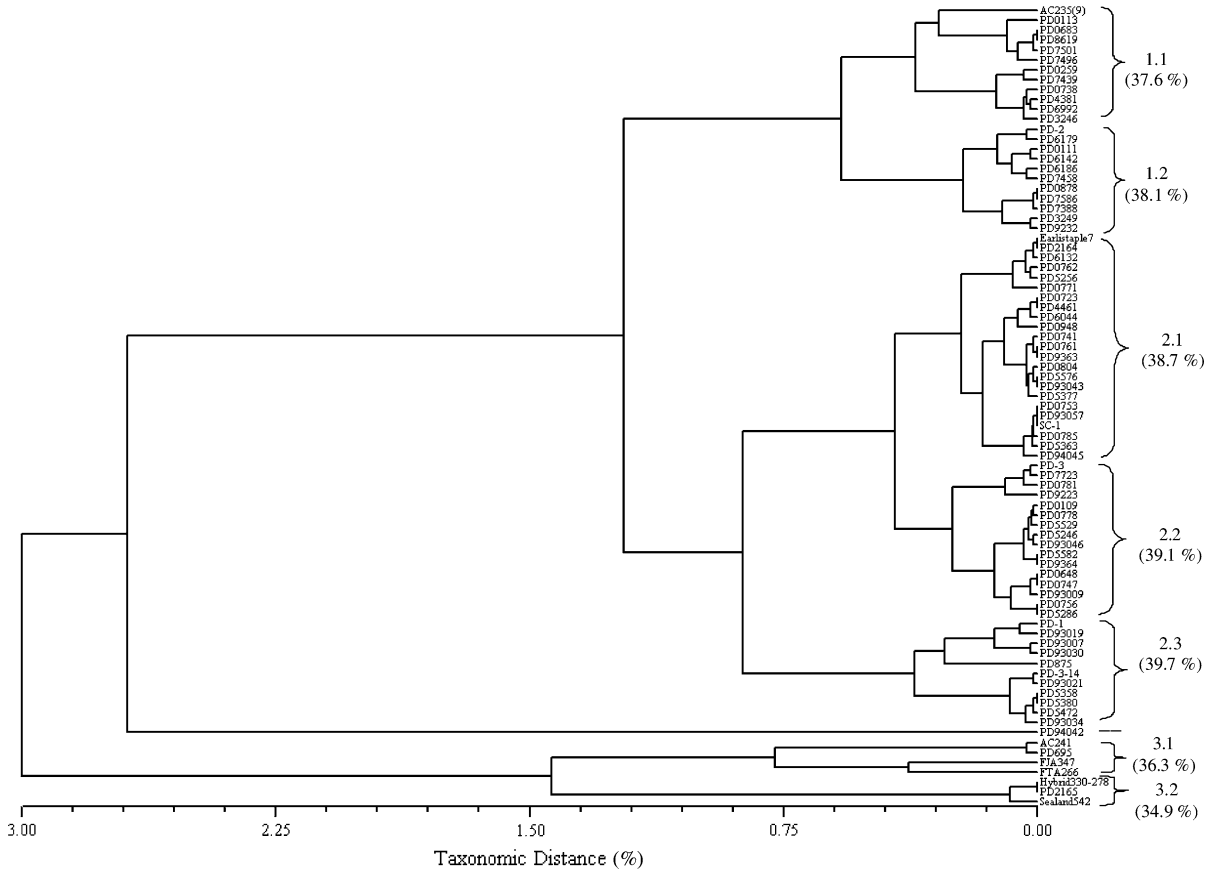

Fig. 2 Dendrogram representing the cluster analysis of Pee Dee germplasm *lines* based on mean lint percent

lint yield sub-groups of primary group 1. Since the early focus of the Pee Dee germplasm program was based on fiber quality introgression (particularly fiber strength), it is not surprising the progenitor lines produced lower yields on average. The pedigree information available for lines clustered in primary group 2 shows that most of these lines were developed more recently from crosses involving a high yield parent on one side of the pedigree.

Figure 4 displays the dendrogram for fiber strength. The Pee Dee lines separated into 2 primary groups with 32 lines clustering to primary group 1 and 49 lines into primary group 2. Primary group 1 and group 2 could each be further separated into 3 sub-groups. The mean fiber strength for primary group 1 was 300 kN m kg^{-1}, while the mean for primary group 2 was 283 kN m kg^{-1}. It was not surprising that each of the early, progenitor Pee Dee lines clustered together into primary group 1 and displayed higher fiber strength on average. High fiber strength was the primary focus of the Pee Dee germplasm program for more than 50 years. Examination of the 3 sub-groups of primary group 1 revealed further separation of this high strength cluster. Sealand542 and Hybrid330-278 further clustered into sub-group 1.3, FTA266 and Earlistaple7 clustered into sub-group 1.1, and FJA347 clustered into sub-group 1.2. Genotypes present in sub-group 1.3 consisted of 5 lines and produced a mean fiber strength of 312 kN m kg^{-1}. Considering the sub-clusters of primary group 1, it was not surprising that most of the Pee Dee lines clustered into sub-group 1.2 with a mean fiber strength of 296 kN m kg^{-1}. The widely known negative relationship between fiber strength and lint yield (Meredith 2005) likely contributes to this clustering pattern of the Pee Dee lines for fiber strength since increased lint yield potential has been a breeding objective in recent years. However, consistent with the low, negative Pearson correlation coefficient between lint yield and fiber strength ($r = -0.08$, $P < 0.05$), our cluster

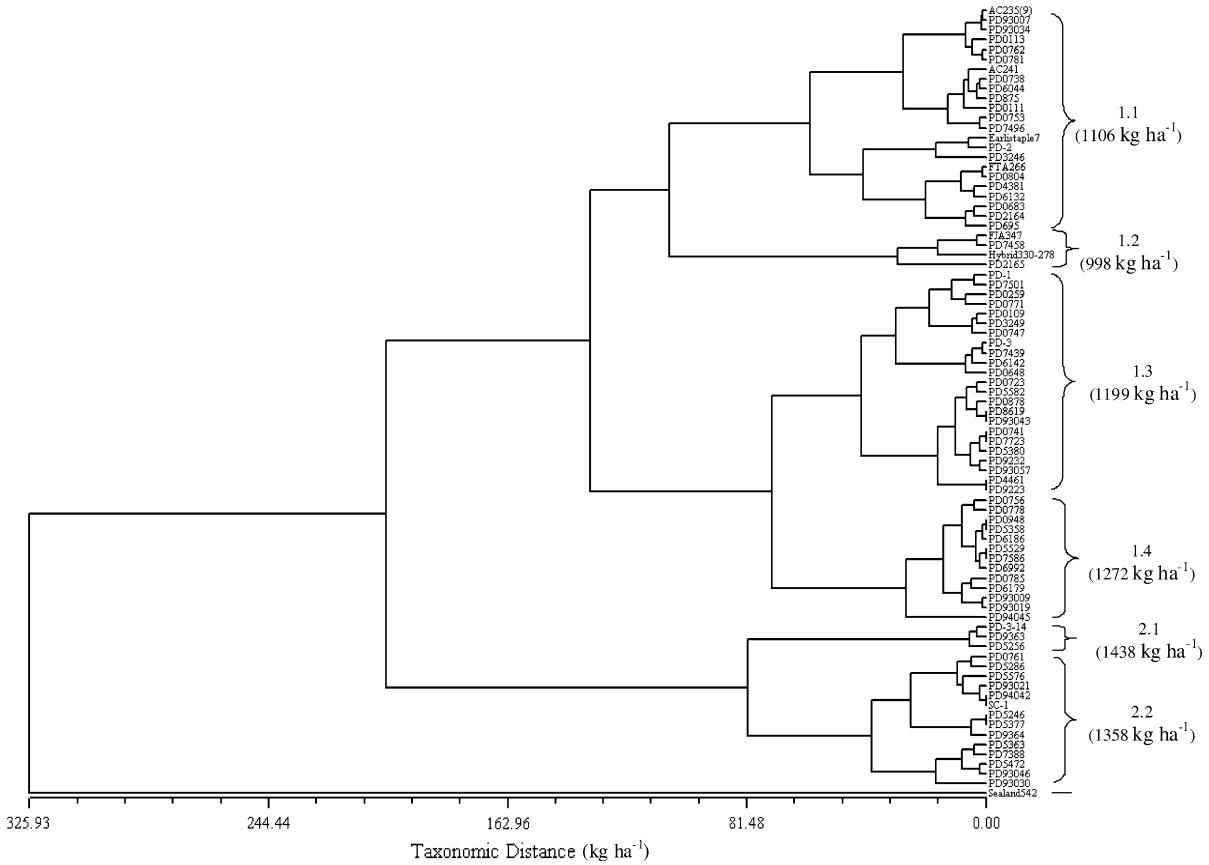

Fig. 3 Dendrogram representing the cluster analysis of Pee Dee germplasm *lines* based on mean lint yield

analysis does reveal several more recently developed Pee Dee lines clustered in sub-groups 1.3 and 1.1 (303 kN m kg^{-1}).

Figure 5 displays the dendrogram for fiber length. The Pee Dee lines separated into 2 primary groups with 9 lines clustering to primary group 1 and 71 lines into primary group 2. Sealand542 did not readily cluster with either of the 2 primary groups. Primary group 2 could be further separated into 4 sub-groups. The mean fiber length for group 1 was 30.0 mm and 28.7 mm for group 2. Similar to the pattern for fiber strength, 3 of the early progenitor lines clustered into group 1. Earlistaple7 did not fall into group 1 but clustered into sub-group 2.2 which displayed a mean fiber length of 29.1 mm. Sealand542 displayed the highest fiber length of any Pee Dee lines and did not readily cluster into either primary group.

Figure 6 displays the dendrogram for fiber fineness. The Pee Dee lines separated into 3 primary groups with 44 lines clustering to primary group 1, 34 lines into primary group 2, and 2 into primary group 3. Sealand542 did not readily cluster with any of the primary groups. Primary groups 1 and 2 could each be further separated into 2 sub-groups. The mean fiber fineness for group 1 was 174.3 mg km^{-1}, 179.4 mg km^{-1} for group 2, and 184.4 mg km^{-1} for group 3. Interestingly, 3 of the early progenitor lines clustered into the finest fiber sub-group 1.2, while only Earlistaple7 clustered into the least fine sub-group 2.1. This indicates the favorable high strength and length properties of FJA347, FTA266, and Hybrid330-278 are likely associated with favorable, very fine fibers. However, Earlistaple7 appears to produce fairly coarse fibers that also have desirable strength and length properties. Overall, this trend for fiber fineness shows that fineness has been decreased somewhat in the Pee Dee germplasm since the initial progenitors, which is consistent with the unfavorable Pearson phenotypic correlation between fiber strength and fiber fineness ($r = 0.40$, $P < 0.05$).

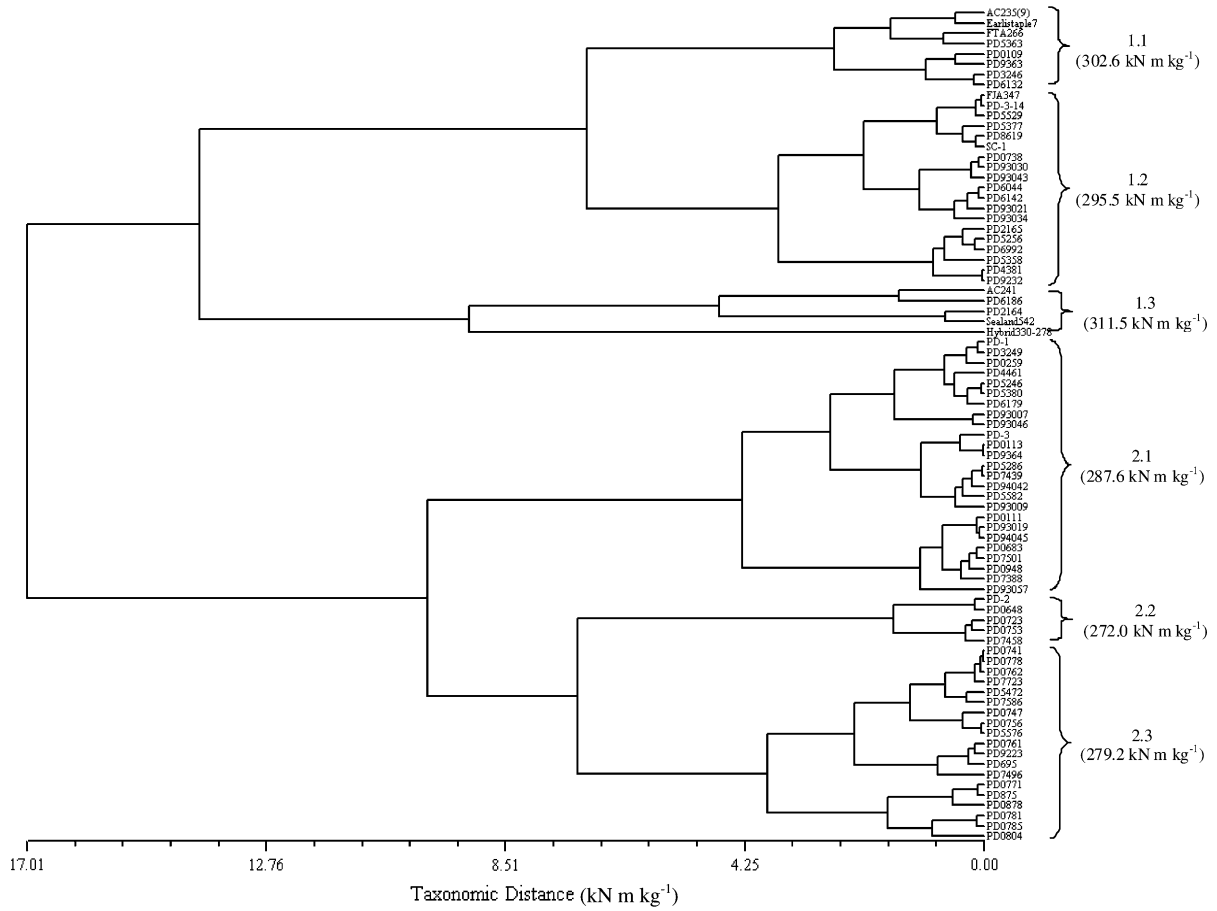

Fig. 4 Dendrogram representing the cluster analysis of Pee Dee germplasm *lines* based on mean fiber strength

Molecular vs. phenotypic trait diversity

When observing the genotype dendrograms present in Figs. 2, 3, 4, 5, and 6, it is clear that genotypes are clustered somewhat differently depending on the dependent variable used to calculate genetic similarity (or dissimilarity) and conduct cluster analysis. We used the normalized Mantel Z statistic (Beer et al. 1993) to test how well the average taxonomic distance matrices calculated for each of the phenotypic traits corresponded to the molecular marker based genetic similarity matrix. In addition, taxonomic distance matrices calculated for each trait were compared with one another. Correlation coefficients for each of the Mantel Z statistic tests are provided in Table 3.

In general, the molecular marker-based genetic similarity matrix did not correspond very closely with the average taxonomic distance matrices for each phenotypic trait. Only fiber length and fiber strength average taxonomic distance matrices were significantly correlated with the molecular marker based genetic similarity matrix. The strongest correlation with molecular markers was found with fiber length ($r = -0.21$). This indicates that genotypes clustered somewhat similarly when based on fiber length phenotypic values and molecular marker values.

Considering the similarity among trait average taxonomic distance matrices, each pairwise combination of traits was highly significant. The correlation coefficients ranged from 0.16 for fiber strength and fineness to 0.57 for lint percent and fiber length. The relative large correlation coefficients between fiber length and fineness (0.41), lint percent and fineness (0.43), fiber length and strength (0.55), and fiber strength and fineness (0.57) indicate these pairwise combinations of traits clustered the Pee Dee genotypes similarly.

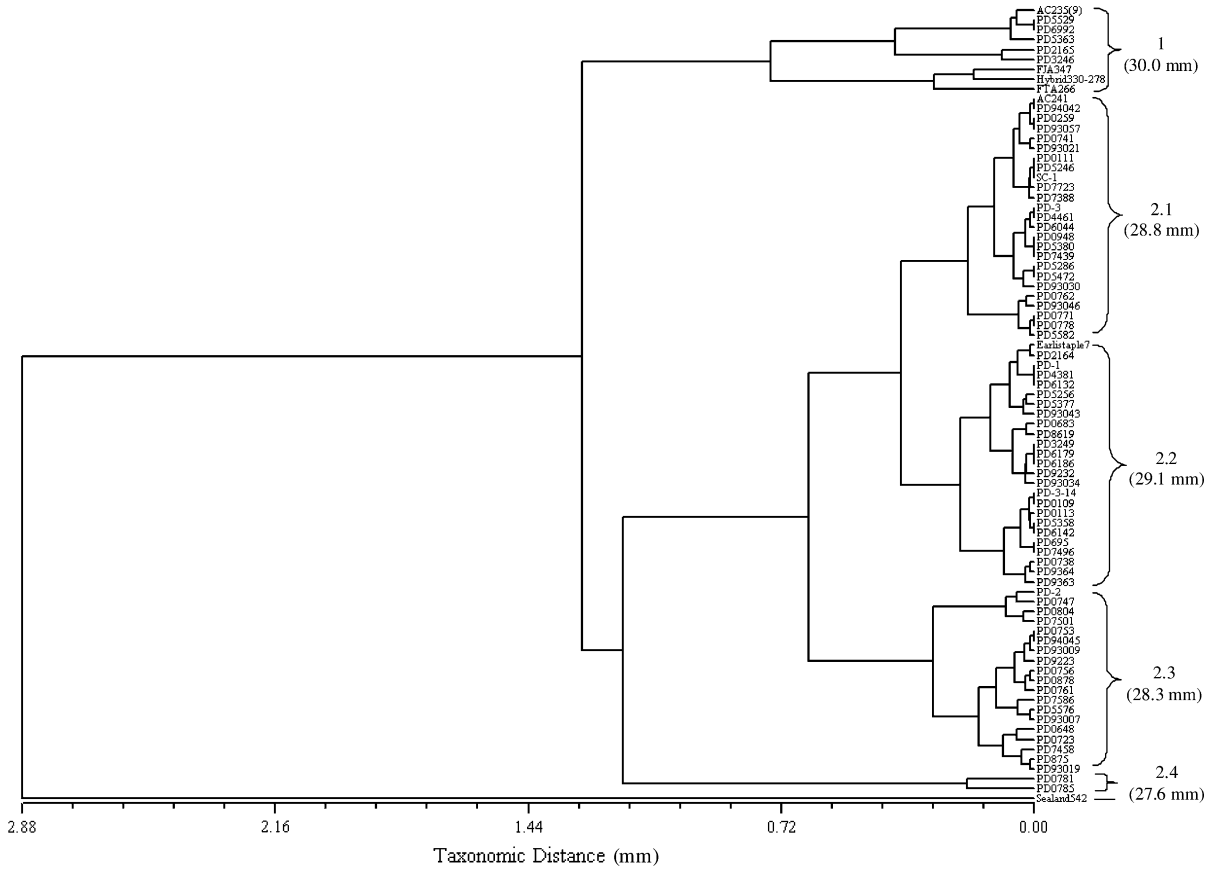

Fig. 5 Dendrogram representing the cluster analysis of Pee Dee germplasm *lines* based on mean fiber length

Conclusion

Results of the molecular marker and phenotypic trait based genetic diversity and cluster analyses suggest that useful genetic diversity remains in the Pee Dee germplasm collection following over 50 years of breeding. Comparing the mean agronomic and fiber quality performance data collected in this study also identifies several Pee Dee lines performing equivalent to or better than commercial cultivars for the traits measured, particularly fiber length, fiber strength, and fiber fineness. The Pearson phenotypic correlations among traits also suggests that progress has been made to lessen the unfavorable relationships involving lint percent, lint yield, and fiber properties. Overall, the degree of molecular marker-based diversity within the Pee Dee germplasm collection is similar to that found by Zhang et al. (2005b) in the New Mexico Acala germplasm program. This finding is not surprising considering the similar, early breeding histories of the Pee Dee and New Mexico Acala germplasm programs and their diverse and unique foundations. The maintenance of Pee Dee germplasm genetic diversity also indicates the alternative breeding methods employed early in the breeding program were successful in facilitating recombination and developing new combinations of alleles. Similar to reports in other crops (Beer et al. 1993), correlations between molecular marker based genetic distance and phenotypic trait taxonomic distance estimates were generally low. However, the molecular marker and phenotypic trait dendrograms generated in this study provide a potential parental line selection tool for plant breeders.

Although this report focused on characterizing and exploiting cotton germplasm, plant breeders of other crops can apply these methods to characterize specific germplasm and generate datasets that can be used as a parental line selection tool for plant breeding programs. Using the specific information provided here, cotton breeders have the ability to select Pee Dee germplasm

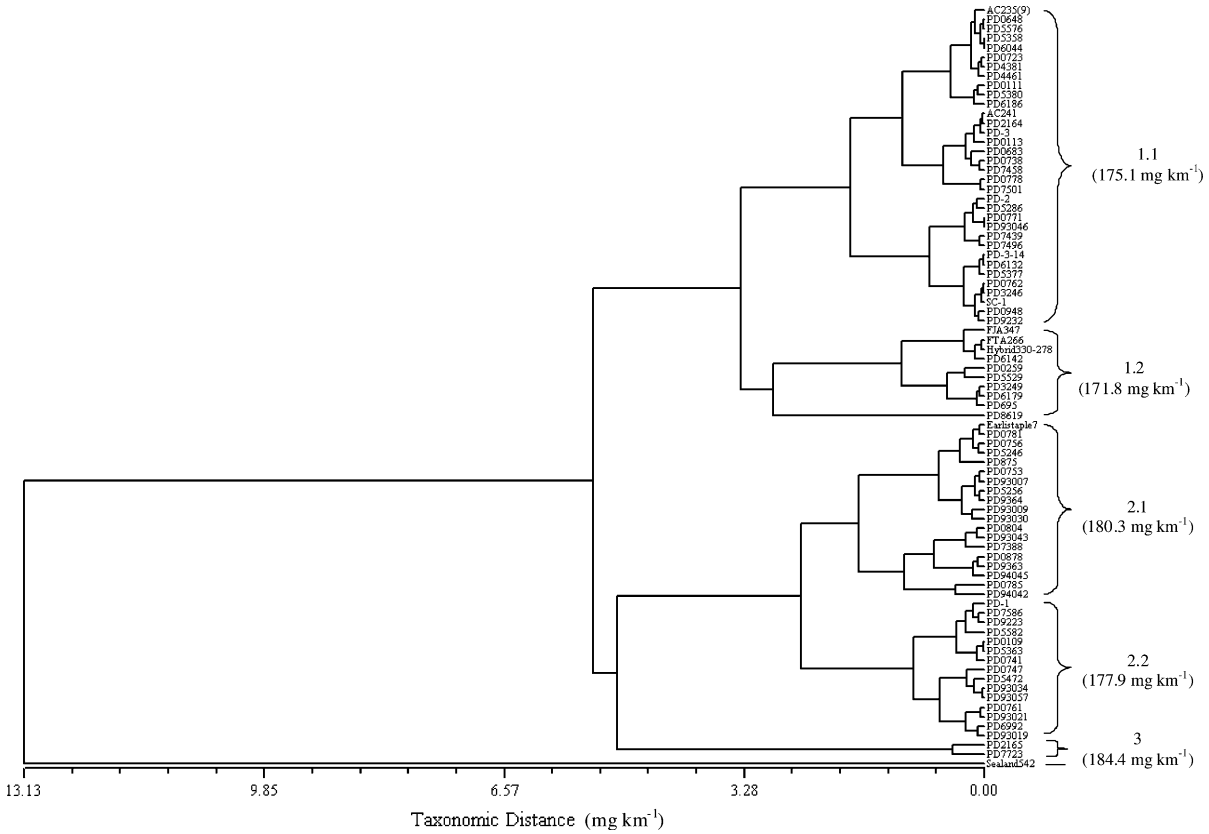

Fig. 6 Dendrogram representing the cluster analysis of Pee Dee germplasm *lines* based on mean fiber fineness

Table 3 Mantel *Z* statistic correlation coefficients among the average taxonomic distance values for lint percent, lint yield, fiber length, fiber strength, and fiber fineness and the molecular genetic similarity matrix

Variable	Lint percent	Lint yield	Fiber length	Fiber strength	Fiber fineness
Molecular markers	−0.03	−0.02	−0.21**	−0.09*	−0.07
Lint percent	–	0.35**	0.57**	0.26**	0.43**
Lint yield	–	–	0.32**	0.18**	0.23**
Fiber length	–	–	–	0.55**	0.41**
Fiber strength	–	–	–	–	0.16**

* Significant at $P < 0.05$

** Significant at $P < 0.01$

parental line combinations based on a combination of molecular and field performance information rather than relying on pedigree and performance data alone. Future research should use these selection criteria to test the hypothesis that hybridizing genetically diverse parents has the highest probability of generating new and favorable allele combinations.

Acknowledgments This research project was supported by funding from CRIS No. 6657-21000-005-00D of the U.S. Department of Agriculture and a grant from Cotton Incorporated. Special thanks to Bobby Fisher and summer students for technical assistance. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

References

Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor Appl Genet 102:222–229. doi:10.1007/s001220051639

Beasley JO (1940) The origin of American *Gossypium* species. Am Nat 74:285–286. doi:10.1086/280895

Beer SC, Goffreda J, Phillips TD, Murphy JP, Sorrells ME (1993) Assessment of genetic variation in *Avena sterilis* using morphological traits, isozymes, and RFLPs. Crop Sci 33:1386–1393

Blenda A, Scheffler J, Scheffler B, Palmer M, Lacape J-M, Yu J, Jesudurai C, Jung S, Muthukumar S, Yellambalase P, Ficklin S, Staton M, Eshelman R, Ulloa M, Saha S, Burr B, Liu S, Zhang T, Fang D, Pepper A, Kumpatla S, Jacobs J, Tomkins J, Cantrell R, Main D (2006) CMD: a cotton microsatellite database resource for *Gossypium* genomics. BMC Genomics 7:132. doi:10.1186/1471-2164-7-132

Bowman DT, Gutierrez OA (2003) Sources of fiber strength in the US upland cotton crop from 1980 to 2000. J Cotton Sci 7:164–169

Bowman DT, Gutierrez OA, Percy RG, Calhoun DS, May OL (2006) Pedigrees of Upland and Pima cotton cultivars released between 1970 and 2005. Bull. 1017. Miss. Agric. and For. Exp. Stn, Mississippi State, MS

Brown-Guedira GL, Thompson JA, Nelson RL, Warburton ML (2000) Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. Crop Sci 40:815–823

Brubaker CL, Wendel JF (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*, Malvaceae) using nuclear restriction fragment length polymorphism (RFLP). Am J Bot 81:1309–1326. doi:10.2307/2445407

Budak H, Pedraza F, Cregan PB, Baenziger PS, Dweikat I (2003) Development and utilization of SSRs to estimate the degree of genetic relationships in a collection of pearl millet germplasm. Crop Sci 43:2284–2290

Culp TW, Harrell DC (1973) Breeding methods for improving yield and fiber quality of upland cotton (*Gossypium hirsutum* L.). Crop Sci 13:686–689

Culp TW, Harrell DC (1979a) Registration of SC-1 cotton. Crop Sci 19:410

Culp TW, Harrell DC (1979b) Registration of five germplasm lines of cotton. Crop Sci 19:751

Culp TW, Harrell DC (1979c) Registration of Pee Dee 4461 cotton germplasm. Crop Sci 19:752

Culp TW, Harrell DC (1979d) Registration of Pee Dee 8619 germplasm line of cotton. Crop Sci 19:753

Culp TW, Harrell DC (1980a) Registration of medium staple cotton germplasm. Crop Sci 20:290

Culp TW, Harrell DC (1980b) Registration of three germplasm lines of cotton. Crop Sci 20:288–289

Culp TW, Harrell DC (1980c) Registration of extra-long staple cotton germplasm. Crop Sci 20:291

Culp TW, Moore RF, Pitner JB (1985a) Registration of PD-1 cotton. Crop Sci 25:198

Culp TW, Moore RF, Pitner JB (1985b) Registration of PD-2 cotton. Crop Sci 25:198–199

Culp TW, Moore RF, Pitner JB (1985c) Registration of seven cotton germplasm lines. Crop Sci 25:201–202

Culp TW, Moore RF, Harvey LH, Pitner JB (1988) Registration of 'PD-3' cotton. Crop Sci 28:190

Culp TW, Green CC, Kittrell BU (1990a) Registration of seven germplasm lines of upland cotton with resistance to bollworm, tobacco budworm, and boll weevil. Crop Sci 30:236–237

Culp TW, Green CC, Kittrell BU (1990b) Registration of twelve noncommercial germplasm lines of upland cotton with resistance to bollworm, tobacco budworm, and boll weevil. Crop Sci 30:236

Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302. doi:10.2307/1932409

Fufa H, Baenziger P, Beecher BS, Dweikat I, Graybosch R, Eskridge K (2005) Comparison of phenotypic and molecular marker-based classifications of hard red winter wheat cultivars. Euphytica 145:133–146. doi:10.1007/s10681-005-0626-3

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638. doi:10.1534/genetics.104.035642

Green CC, Culp TW, Kittrell BU (1991a) Registration of four germplasm lines of upland cotton with early maturity and high fiber quality. Crop Sci 31:854

Green CC, Culp TW, Kittrell BU (1991b) Registration of five germplasm lines of upland cotton with high yield potential and fiber quality. Crop Sci 31:854–855

Green CC, Culp TW, Kittrell BU (1991c) Registration of two germplasm lines of upland cotton with highyield potential and fiber quality. Crop Sci 31:853

Gutierrez OA, Basu S, Saha S, Jenkins JN, Shoemaker DB, Cheatham CL, McCarty JC Jr (2002) Genetic distance among selected cotton genotypes and its relationship with F2 performance. Crop Sci 42:1841–1847

Hanson WD (1959) The breakup of initial linkage blocks under selected mating systems. Genetics 44:857–868

Harrell DC, Culp TW (1979a) Registration of Pee Dee 0259 and Pee Dee 2165 germplasm lines of cotton. Crop Sci 19:418

Harrell DC, Culp TW (1979b) Registration of Pee Dee 4381 germplasm line of cotton. Crop Sci 19:418

Institute SAS (2008) The SAS system for Windows, Release 9.2. SAS Inst, Cary, NC

Iqbal J, Reddy OUK, El-Zik KM, Pepper AE (2001) A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L examined using DNA fingerprinting. Theor Appl Genet 103:547–554. doi:10.1007/PL00002908

Jiang C, Chee PW, Draye X, Morrell PL, Smith CW, Paterson AH (2000) Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). Evol Int J Org Evol 54:798–814

Kerr T (1960) The trispecies hybrid ancestry of high strength cottons. Proc Cotton Imp Conf 21:82

Khan SA, Hussain D, Askari E, Stewart JM, Malik KA, Zafar Y (2000) Molecular phylogeny of *Gossypium* species by DNA fingerprinting. Theor Appl Genet 101:931–938. doi:10.1007/s001220051564

Labate JA, Lamkey KR, Mitchell SE, Kresovich S, Sullivan H, Smith JSC (2003) Molecular and historical aspects of corn belt Dent diversity. Crop Sci 43:80–91

Lacape J-M, Dessauw D, Rajab M, Noyer J-L, Hau B (2007) Microsatellite diversity in tetraploid *Gossypium* germplasm:

assembling a highly informative genotyping set of cotton SSRs. Mol Breed 19:45–58. doi:10.1007/s11032-006-9042-1

Liu S, Cantrell RG, McCarty JC Jr, Stewart JM (2000) Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. Crop Sci 40:1459–1469

Lu HJ, Myers GO (2002) Genetic relationships and discrimination of ten influential Upland cotton varieties using RAPD markers. Theor Appl Genet 105:325–331. doi:10.1007/s00122-002-0947-8

May OL (1999) Registration of PD 94042 germplasm line of upland cotton with high yield and fiber maturity. Crop Sci 39:597–598

May OL (2001) Registration of PD 94045 germplasm line of upland cotton. Crop Sci 41:279–280

May OL, Howle DS (1997a) Registration of six germplasm lines of upland cotton: PD 93009, PD 93019, PD 93021, PD 93030, PD 93034, and PD 93057. Crop Sci 37:1030–1031

May OL, Howle DS (1997b) Registration of three germplasm lines of upland cotton: PD 93007, PD 93043, and PD 93046. Crop Sci 37:1030

May OL, Bowman DT, Calhoun DS (1995) Genetic diversity of U.S. upland cotton cultivars releases between 1980 and 1990. Crop Sci 35:1570–1574

May OL, Howle DS, Green CC, Culp TW (1996) Registration of PD-3-14 germplasm line of upland cotton with high yield and fiber quality. Crop Sci 36:1718

Meredith WR Jr (2005) Minimum number of genes controlling cotton fiber strength in a backcross population. Crop Sci 45:1114–1119. doi:10.2135/cropsci2003.0425

Milla SR, Isleilb TR, Stalker HT (2005) Taxonomic relationships among *Arachis* sect *Arachis* species as revealed by AFLP markers. Genome 48:1–11. doi:10.1139/g04-089

Mohammadi SA, Prasanna BM (2003) Analysis of genetic diversity in crop plants-salient statistical tools and considerations. Crop Sci 43:1235–1248

Nei M, Li W (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273. doi:10.1073/pnas.76.10.5269

Nguyen T-B, Giband M, Brottier P, Risterucci A-M, Lacape J-M (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. Theor Appl Genet 109:167–175. doi:10.1007/s00122-004-1612-1

Pillay M, Myers GO (1999) Genetic diversity in cotton assessed by variation in ribosomal RNA genes and AFLP markers. Crop Sci 39:1881–1886

Rahman M, Hussain D, Zafar Y (2002) Estimation of genetic divergence among elite cotton cultivars-genotypes by DNA fingerprinting technology. Crop Sci 42:2137–2144

Rohlf FJ (2005) NTSYSpc, numerical taxonomy and multivariate analysis system, version 2.2. Applied Biostatistics Inc, New York

Smith CW, Cantrell RG, Moser HS, Oakley SR (1999) History of cultivar development in the United States. In: Smith CW, Cothren JT (eds) Cotton: origin, history, technology, and production. John Wiley & Sons Inc, New York, pp 99–171

Tatineni V, Cantrell RG, Davis DD (1996) Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. Crop Sci 36:186–192

Van Becelaere G, Lubbers EL, Paterson AH, Chee PW (2005) Pedigree- vs DNA marker-based genetic similarity estimates in cotton. Crop Sci 45:2281–2287. doi:10.2135/cropsci2004. 0715

Van Esbroeck G, Bowman DT, Calhoun DS, May OL (1998) Changes in the genetic diversity of cotton in the USA from 1970 to 1995. Crop Sci 38:33–37

Van Esbroeck G, Bowman D, May O, Calhoun D (1999) Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. Crop Sci 39:323–328

Wang K, Song X, Han Z, Guo W, Yu J, Sun J, Pan J, Kohel R, Zhang T (2006) Complete assignment of the chromosomes of *Gossypium hirsutum* L by translocation and fluorescence in situ hybridization mapping. Theor Appl Genet 113:73–80. doi:10.1007/s00122-006-0273-7

Wendel JF, Brubaker CL, Percival AE (1992) Genetic diversity in *Gossypium hirsutum* and the origin of Upland cotton. Am J Bot 79:1291–1310. doi:10.2307/2445058

Zhang J, Lu Y, Cantrell RG, Hughs E (2005a) Molecular marker diversity and field performance in commercial cotton cultivars evaluated in the southwestern USA. Crop Sci 45:1483–1490. doi:10.2135/cropsci2004.0581

Zhang JF, Lu Y, Adragna H, Hughs E (2005b) Genetic improvement of New Mexico Acala cotton germplasm and their genetic diversity. Crop Sci 45:2363–2373. doi:10.2135/cropsci2005.0140